# The Instantiation and Use of Conceptual Simulations: Movies in the Mind to Connect Theory and Data in Scientific Visualization


**Susan B. Trickett**
**(stricket@gmu.edu)**
Department of Psychology,
George Mason University
Fairfax, VA 22030 USA

**J. Gregory Trafton**
**(trafton@itd.nrl.navy.mil)**
Naval Research Laboratory
NRL Code 5513
Washington, DC 20375

**Christian D. Schunn**
**(schunn@gmu.edu)**
Department of Psychology,
George Mason University
Fairfax, VA 22030 USA



## Abstract

We investigate the relationship among internal and external visualizations and domain knowledge in expert scientists. We observed scientists as they analyzed their own data using computer-based visualization tools. All the scientists used sequences of dynamic mental images (conceptual simulations) in their investigations. They used conceptual simulations primarily to evaluate hypotheses about the appearance of the data. They measured the result of the simulation against the empirical data in a process of alignment that allowed them to estimate the fit between the hypothesized and the actual results. A model of this process is discussed.


## Introduction

In his most famous "thought experiment," Einstein imagined himself traveling next to a beam of light. Insights from this mental visualization process led eventually to his formulation of the special theory of relativity (Shepard, 1988). According to Einstein himself, the use of "more or less clear images which can be 'voluntarily' reproduced and combined" was crucial to his thinking (Hadamard, 1945).

The use of visual imagery, whether internal or external, appears to be an important aspect of science. Shephard (1988) has identified numerous instances of the importance of mental visualization in scientific discovery and problem solving for many famous scientists. Furthermore, external visualization is used by almost all scientists, at the very least to represent and display data as an aid to analysis, if not as an aid to scientific discovery. In contemporary science, these visualizations are frequently computer-generated and may range from relatively simple 2-dimensional graphs to highly complex representations of multi-dimensional data. The advantages of using external graphical representations have been demonstrated by, for example, Cheng and Simon (1995), Larkin and Simon, (1987), and Tabachneck-Schijf, Leonardo, and Simon, (1997).

Although external visualizations are commonly used by scientists to explore large datasets, little is known about *how* complex visualizations are actually used or how they fit into the important cycle of hypothesis generation and evaluation that constitutes a crucial component of scientific reasoning. However, it does seem to be the case that the experts' use of visualizations involves mentally manipulating displayed images in order to extract additional, implicit information from them (Trafton et al., 2000). It also appears that people can and do apply this kind of manipulation to ad hoc mental images that they have created to solve a particular problem (Qin & Simon, 1990). Trafton has proposed a framework of spatial transformations by which these manipulations can be categorized and investigated (Trafton, Trickett, & Mintz, 2001).

In addition to strategic knowledge (such as data interpretation skills), an important component of scientific thinking is the extensive domain knowledge that scientists acquire during their years of training (Ericsson & Charness, 1994; Schunn & Anderson, 1999). As scientists inspect visualizations of their data, their domain knowledge not only guides them to look for certain patterns but also provides a basis for explaining those patterns. How, then, do scientists integrate their theoretical (domain) knowledge with data in these external representations? When they generate hypotheses to account for observed data, how do they evaluate those hypotheses? What role, if any, is played by processes of internal visualization, such as those described by Einstein?

The goal of this paper is to investigate the relationship among scientists' theoretical domain knowledge, the external data visualizations they use, and their use of internal, or mental, visualization. We propose that scientists bring their theoretical knowledge to bear on currently displayed data through a specific process of mental visualization that we term "conceptual simulation." We argue that they create this conceptual simulation based upon their domain knowledge or on the current hypothesis. We suggest that they first create a set of dynamic mental images, which acts as a sort of "movie in the mind" (conceptual simulation), and then overlay the end product of this simulation on the actual data represented in the visualization (a process of alignment). A close match indicates support for the accuracy of the simulation, whereas discrepancies are viewed as evidence that the simulation cannot account for the empirical data. Finally, we propose that this sequence of cognitive operations is primarily a strategy for the evaluation of hypotheses.

## Method

In order to investigate the issues discussed above, we have adapted Dunbar's in vivo methodology (Dunbar, 1995; Trickett, Trafton & Schunn, 2000). This approach offers several advantages. First, it allows observation of experts, who can use their domain knowledge to guide their strategy selection. Second, it allows the collection of "on-line" measures of thinking, so that the scientists' reasoning can be examined as it occurs (Ericsson & Simon, 1993). Finally, the tasks (experiment design, data analysis, etc) conducted by the scientists, as well as the tools they use, are fully authentic.

Two sets of scientists were videotaped while conducting their own research. All the scientists were experts, having earned their Ph.D.s more than 6 years previously. In the first set, two astronomers, one a tenured professor at a university, the other a fellow at a research institute, worked collaboratively to investigate computer-generated visual representations of a new set of observational data. At the time of this study, one astronomer had approximately 20 publications in this general area, and the other approximately 10. The astronomers have been collaborating for some years, although they do not frequently work at the same computer screen and the same time to examine data.

In the second dataset, a physicist with expertise in computational fluid dynamics worked alone to inspect the results of a computational model he had built and run. He works as a research scientist at a major U.S. scientific research facility, and had earned his Ph.D. 23 years ago. Having inspected the data earlier, had made some adjustments to the physics parameters underlying the model and was therefore revisiting the data.

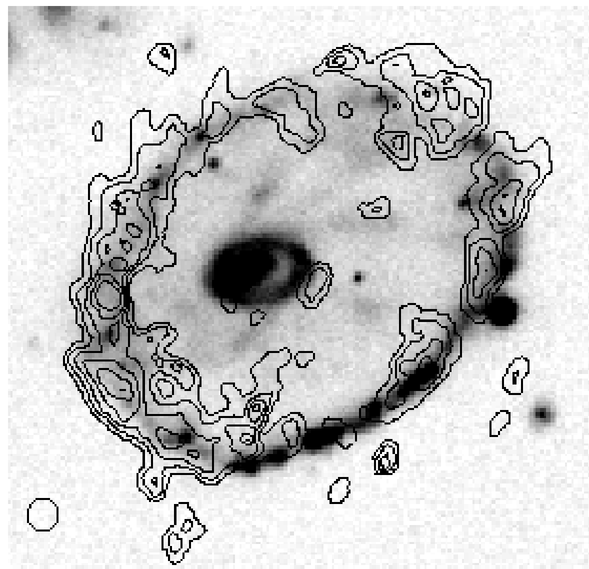
Both sets of scientists were instructed to carry out their work as though no camera were present, without explanation to the experimenter (Ericsson & Simon, 1993). The relevant part of the astronomy session lasted about 53 minutes, and the physics session, 15 minutes. All utterances were transcribed and segmented according to complete thought. All segments were coded by 2 coders as on-task (pertaining to data analysis) or off-task (e.g., jokes, phone interruptions, etc.). Inter-rater reliability for this coding was over 95%. Off-task segments were excluded from further analysis. On-task segments (N = 649 [astronomy] and N = 176 [physics]) were further coded as described below.

### The Tasks and the Data

**Astronomy** The data under analysis were optical and radio data of a ring galaxy. The astronomers' high-level goal was to understand its evolution and structure by understanding the flow of gas in the galaxy. In order to understand the gas flow, the astronomers must make inferences about the velocity field, represented by contour lines on the 2-dimensional display.

The astronomers' task was made difficult by two characteristics of their data. First, the data were one- or at best 2-dimensional, whereas the structure they were attempting to understand was 3-dimensional. Second, the data were noisy, with no easy way to separate noise from real phenomena. Figure 1 shows a screen snapshot of the type of data the astronomers were examining. In order to make their inferences, the astronomers used different types of image, representing different phenomena (e.g., different forms of gas), which contain different information about the structure and dynamics of the galaxy. In addition, they could choose from images created by different processing algorithms, each with advantages and disadvantages (e.g., more or less resolution). Finally, they could adjust some features of the display, such as contrast or false color.

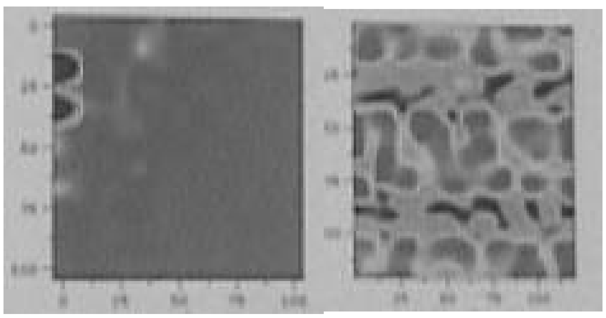
Figure 1: Example of data examined by astronomers. Radio data (contour lines) are laid over optical data.



**Physics** The physicist was working to evaluate how deep into a pellet a laser light will go before being reflected. His high-level goal was to understand the fundamental physics underlying the reaction, an understanding that hinged on an understanding of the relative importance and growth rates of different modes. The physicist had built a model of the reaction; other scientists had independently conducted experiments in which lasers were fired at pellets and the reactions recorded. A close match between model and empirical data would indicate a good understanding of the underlying theory. Although the physicist had been in conversation with the experimentalist, he had not viewed the empirical data, and in this session he was investigating only the results of his computational model. However, he believed the model to be correct (i.e., he had strong expectations about what he would see), and in this sense, this session may be considered confirmatory.

The data consisted of two different kinds of representation of the different modes, shown over time (nanoseconds). The physicist was able to view either a Fourier decomposition of the modes or a representation of the “raw” data. Figure 2 shows an example of the physicist's data. He could choose from black-and-white or a variety of color representations, and could adjust the scales of the displayed image, as well as some other features. He was able to open numerous views simultaneously. A large part of his task was comparing images, both different types of representation of the same data and different time slices represented in the same way.

Figure 2: Example of data examined by physicist Fourier modes (left) and raw data (right)



**Coding Scheme**

Our goals in this research are first to establish the existence of conceptual simulations across different datasets and second, to investigate their use in the analysis of data by expert scientists using their own data visualization tools. We propose that conceptual simulations are used to evaluate hypotheses, and that this evaluation occurs when the scientists align the results of the simulation against the actual data. Consequently, we identified all hypotheses proposed by the scientists, all conceptual simulations, and all utterances that align a mental image with the data in the current visualization.

These processes involve cognitive operations on both internal and external images. Trafton's spatial transformation framework (Trafton, Trickett, & Mintz, 2001) is designed to capture mental manipulations of images. Therefore, in order to maximize the reliability with which these codes were applied, we used a Spatial Transformation Analysis (explained below) to identify conceptual simulations and alignments in each protocol. We now describe this coding scheme (hypothesis, conceptual simulation, alignment) in detail.

**Hypotheses** Statements that attempted to account for the appearance of the data were coded as hypotheses. For example Astronomer 1 (hereafter referred to as A1) noticed an area where the velocity contour "sort of dips under, sort of does a very non-circular motion thing." Astronomer 2 (hereafter referred to as A2) asked if this is "significant." A1 suggested that the phenomenon might indicate a "streaming motion" (hypothesis).

We foresaw that some hypotheses would be dismissed immediately and would therefore not be relevant to our analyses. We identified all hypotheses and coded them as elaborated or unelaborated. Elaboration consisted of one or more statements supporting or opposing the hypothesis. Those that were not discussed after being proposed were coded as unelaborated.

**Spatial Transformation Analysis** Spatial transformations are cognitive operations that a scientist performs on a visualization. Sample spatial transformations are mental rotation (Shepard, 1971), creating a mental image, modifying that mental image by adding or deleting features, mentally moving an object, comparisons between different views (Kosslyn, Sukel, & Bly, 1999; Trafton, Trickett, & Mintz, 2001), and anything else a scientist mentally does to a visualization in order to understand it or facilitate problem-solving. Note that a spatial transformation can be done on either an internal (i.e., mental) image or an external image (e.g., a computer-generated visualization). What all spatial transformations have in common is that they involve the use of mental imagery. Statements by which the scientists directly extracted information from the visualization were *not* considered spatial transformations. A more complete description can be found at http://iota.gmu.edu/users/trafton/405st.html.

For every utterance in each protocol we evaluated whether there was a spatial transformation. Spatial transformations were further coded as Create Image, Modify Image (Add or Delete), or Comparison. Table 1 shows examples of each category of spatial transformation (note that these utterances are independent of one another and do not represent a sequence).

| Utterance | Spatial Transformation |
|---|---|
| I can easily imagine a gas as being…. | Create Image |
| The Kittemandefax would be right along this region [points to area on displayed image] | Modify Image: Add |
| If there was no streaming motion or sort of piling of gas | Modify Image: Delete |
| Also, look at the rest of the ring—you see similar kinds of sort of intrusions | Comparison |

Table 1: Examples of Spatial Transformations.

**Alignment** Alignment is a specific type of comparison (Trafton, Trickett, & Mintz, 2001) in which a mental image is overlaid over a displayed image, in order to make an estimation of “fit” between the two images. Statements which compared a built-up mental image with the currently displayed visualization were coded as alignments. For example, at one point, A1 commented, "If you looked at a spiral arm [of a galaxy]…if there was no streaming motion…the lines would show no deflection as they went across the spiral arm." A1 gestures to lines on the current display, indicating that he is comparing these lines with the lines in his mental image of the spiral arm (alignment). In this case, there is a discrepancy between the two images, because the lines in the displayed image *do* show deflection.

**Conceptual Simulations** A conceptual simulation is a mentally constructed model of a phenomenon or data representation. The assumption is that the scientist can always view data in the display, but that he makes use of additional information in memory that is *not* represented on the display, and that this additional information is represented as a mental image (Kosslyn, 1990). The initial image upon which the conceptual simulation is based may be grounded in domain knowledge or in a modification of the displayed image, or may be recalled from memory. They key feature of a conceptual simulation is that it involves a simulation "run" that alters the image; this run may be explicit or implicit, as defined below. Table 2 provides examples of the sequence of spatial transformations that comprises a conceptual simulation, with both explicit and implicit runs.

Conceptual simulations are defined formally as a specific sequence of spatial transformations:

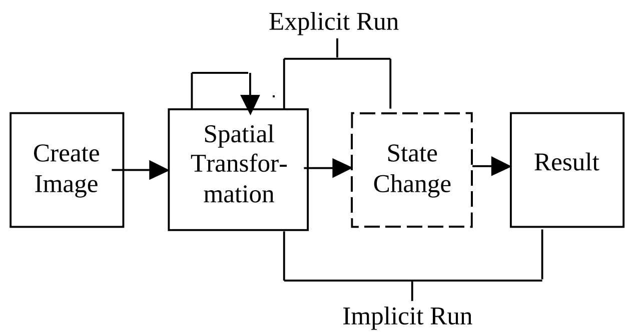
1. Create image: The scientist creates a mental image that is different from the currently displayed image. This image creation may occur via a (mental) modification of the display, via domain knowledge, or via memory (recall of a previously viewed image).

2a). Explicit Run: The scientist builds on the created image by spatial transformation (e.g., extend, add, etc.) such that its state is changed; or

2b). Implicit Run: The scientist refers to a "result" (inspection) of a modified image, which implies the change of state necessary for a run to have occurred.

It should be noted that these codes are not mutually exclusive—the created image and explicit run can occur in the same utterance (though this occurred less than 15% of the time). Figure 3 shows a schematic diagram of a conceptual simulation, highlighting the difference between explicit and implicit runs.

Figure 3: Schematic diagram of a conceptual simulation




## Results

First, we establish the existence of conceptual simulations; we then examine the relationship among conceptual simulations, hypotheses, and alignment.

A subset of the astronomy protocol was coded by two independent coders. Initial inter-rater reliability was more than 90% and all disagreements were resolved by discussion. Because we found high agreement in the as-

| Utterance | Spatial Transform-ation | Explanation |
|---|---|---|
| A2: I mean,… in a perfect sort of spider diagram, | Create | A2 creates image of theoretical spider diagram (not visible on display) |
| if you looked at the velocity contours without any sort of streaming motions, without streaming motions,… | Modify: delete | A2 modifies image of spider diagram by deleting streaming motions, thus changing the state of the phenomenon (explicit run) |
| you'd probably expect these lines here to go all the way across, you know, the ring, without any sort of, um, changes here in the slope and stuff | Align | A2 aligns modified image of spider diagram with actual displayed image and compares the appearance (direction) of the lines. The images do not match, suggesting the presence of streaming motions. |
| A1: Maybe it's a projection effect. | | Hypothesis stated by A1 |
| A2: It's a projection effect, | | Reiteration of hypothesis by A2 |
| A2: although if that's true | Create | A2 creates image in which phenomenon under consideration is, in fact, a "projection effect" |
| A2: there should be a very large velocity dispersion here. | Modify: add Align | A2 modifies image with addition of expected large velocity dispersion. He inspects the result of this modification (implicit run) and aligns to currently displayed image ("here"). The match fails. |
| A2: I don't recall, I don't think I saw anything with velocity dispersion in it | Continued alignment | A2 searches memory for a match but fails to produce one. |

Table 2: Examples of conceptual simulations. A1 and A2 represent astronomer 1 and 2, respectively.

tronomy dataset, we expect agreement in the physics dataset to be high also.

### Do Conceptual Simulations Exist?

In the astronomy dataset, there were 22 conceptual simulations, 11 by each astronomer (approximately one every 2.5 minutes), accounting for 9% of 649 on-task segments. In the physics dataset, there were 4 conceptual simulations (approximately one every 3.75 minutes), accounting for 9% of 176 on-task utterances.

### How Were Conceptual Simulations Used?

In order to determine how conceptual simulations were used, we analyzed the relationship between conceptual simulations and 2 other types of utterance: hypotheses and alignments. Table 3 summarizes these results.

| | Astronomy | Physics |
|---|---|---|
| Hypotheses with conceptual simulations | 76% | 50% |
| Conceptual simulations with hyotheses | 91% | 75% |
| Conceptual simulations with alignment | 86% | 75% |

Table 3: Relationship between conceptual simulations, elaborated hypotheses, and alignment

**Conceptual Simulations and Hypotheses** There were 21 (astronomy) and 9 (physics) hypotheses. Only elaborated hypotheses are considered in the following analyses (81% for astronomy, 66% for physics). As Table 3 shows, the evidence for or against the majority of these hypotheses contained a conceptual simulation.

In addition to the fact that most hypotheses were associated with conceptual simulations, conversely, most of the conceptual simulations were associated with a hypothesis (see Table 3). By far the most frequent use of conceptual simulations was thus in elaborating—supporting or opposing—a hypothesis.

If a conceptual simulation was not associated with a hypothesis, how was it used? In the astronomy dataset, the remaining conceptual simulations were used to clarify theoretical issues necessary to an understanding of the data. For example, A2 uses a conceptual simulation to help understand what happens theoretically in a galaxy with an expanding ring. However, he is not exploring a specific hypothesis at this point, but building a theoretical picture that he then matches against the displayed data. In the physics dataset, only one conceptual simulation did not follow a hypothesis. In this case, the conceptual simulation was used to *develop* a hypothesis; that is, the hypothesis was the outcome of the conceptual simulation. Although there are few instances where conceptual simulations do not elaborate a hypothesis, it does appear that they are used to link theory and data in ways other than evaluating a hypothesis.

**Conceptual Simulations and Alignment** As Table 3 shows, for most of the conceptual simulations, the very next utterance was an alignment which matched the simulation results with the currently displayed image.

**Hypotheses, Conceptual Simulations, and Alignment** The results discussed above suggest a sequence of activity by which the scientists investigated their data, in which a hypothesis was proposed, and then evidence in support of or against this hypothesis was considered. A significant part of this evidence consisted of a conceptual simulation, the results of which were inspected and aligned with the actual data. The extent to which a match was found was taken as evidence for or against the original hypothesis.

## General Discussion and Conclusion

In this research, we have explored the use of internal and external visualizations and their relationship to domain knowledge among expert scientists in two domains. We have shown that these scientists not only use internal, mental visualizations, but also modify and otherwise manipulate those visualizations in conceptual simulations. The fact that these conceptual simulations occur in two quite different datasets indicates that they are not domain specific. Furthermore, their use by three separate individuals suggests that they are not the result of an individual difference. Finally, their use in different situations (by a single scientist working alone, as well as by a dyad) indicates that conceptual simulations are not necessarily a rhetorical device, used to persuade another person of the validity of an argument.

We have further shown that these scientists use these conceptual simulations primarily to evaluate hypotheses when they are analyzing data, and this evaluation takes the form of aligning the results of the simulation with the actual data. The scientists determine how well the outcome of the simulation fits the data in order to assess the validity of the hypothesis on which the simulation rests. This process of alignment and judgment of fit is an important part of the hypothesis-evaluation process.

In this way, conceptual simulations may serve a similar function to full computational models that are increasingly used as a means of hypothesis evaluation in many sciences. Like a conceptual simulation, a computational model also represents the instantiation of the scientist's theoretical assumptions. Once the model is built, it is usually aligned against some empirical data, and measures of fit (formal or informal) are taken. A good fit between model and data is generally accepted as evidence for the validity of the model's underlying assumptions. A poor fit, on the other hand, indicates that those assumptions are incorrect.

However, a computational model is an expensive undertaking (in time and other resources). Conceptual simulations may serve as a kind of preliminary screening strategy that allows the scientist to "weed out" hypotheses that hold no promise of success and focus on those that indicate a good fit with the data. There is some evidence for this interpretation, as A1 comments that the analysis session has been helpful because “I can see where I need to make some models.” Conceptual simulations may thus be considered a type of “modeling-on-the-fly” that offers an inexpensive but effective first pass at evaluating the scientist's conceptual or theoretical understanding of the data.

Our results have several implications. At an applied level, researchers in science education have noted that students have difficulty applying theoretical knowledge —for example, predicting the trajectory of a moving object (e.g., Caramazza, McCloskey, & Green, 1981). One solution has been to build visualization tools, such as computer simulations and virtual reality worlds, that allow students to manipulate visual representations of abstract phenomena and match the results of those manipulations against their predictions (e.g., White, 1993). Our analyses show that this is an authentic strategy used by practicing scientists, although the scientists, of course, conduct these manipulations mentally.

At a theoretical level, current models of scientific reasoning do not include the type of mental manipulation of imagery represented by conceptual simulations; however, it appears that this is an important component of the process of data analysis, and therefore of “doing science.” A similar set of processes may underlie the function-follows-form reasoning packets identified by Griffith, Nersessian, & Goel (2000). The relationship between conceptual simulations and other model-based reasoning strategies warrants further investigation.

## Acknowledgments

This research was supported in part by grant 55-7850-00 to the second author from ONR. We thank Melanie Diez, Wai-Tat Fu, Anthony Harrison, William Liles, and Lelyn Saner for their comments.

## References

Carramazza, A., McCloskey, M., & Green, B. (1981). Naïve beliefs in "sophisticated" subjects: Misconceptions about trajectories of objects. *Cognition, 9*(2), 117-123.

Cheng, P. C-H., & Simon, H. A. (1995). Scientific ciscovery and creative reasoning with diagrams. In S. M. Smith, T. B. Ward, & R. A. Finke, Eds., *The creative cognition approach* (pp. 203-228). Cambridge, MA: MIT Press.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson, Eds., *The nature of insight* (pp. 365-395). Cambridge, MA: MIT Press.

Ericsson, K. A., & Charness, (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*(8), 725-747.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data.* 2nd edition. Cambridge, MA: MIT Press.

Griffith, T. W., Nersessian, N. J., & Goel, A. (2000). Function-follows-form transformations in scientific problem solving. In *The Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 196-201). Hillsdale, NJ: Erlbaum.

Hadamard, J. (1945). *The psychology of invention in the mathematical field*. Princeton, NJ: Princeton.

Kosslyn, S. M. (1990). Mental imagery. In D. N. Osherson, S. M. Kosslyn, & J. M. Hollerback, Eds., *Visual cognition and action: An invitation to cognitive science, Vol.2* (pp. 73-97). Cambridge, MA: MIT Press.

Kosslyn, S. M., Sukel, K. E., & Bly, B. M. (1999). Squinting with the mind's eye: Effects of stimulus resolution on imaginal and perceptual comparison. *Memory & Cognition, 27*(2), 276-287.

Larkin, J. H. & Simon, H. A. (1987). Why a diagram is (sometimes) worth 10,000 words. *Cognitive Science, 4*, 317-345.

Qin, Y. & Simon, H. A. (1990). Imagery and problem solving. In *The Proceedings of the 12th Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Schunn, C. D., & Anderson, J. R. (1999). The generalitiy/specificity of expertise in scientific reasoning. *Cognitive Science, 23*(3), 337-370.

Shepard, R. (1988). The imagination of the scientist. in E. Egan & D. Nadaner, Eds., *Imagination and education* (pp. 153-185). New York, NY: Teachers College Press.

Shepard, R. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 00. 701-703.

Tabachneck-Schijf, H. J. M., Leonardo, A. M., & Simon, H. A. (1997) CaMeRa: A computational model of multiple representations. *Cognitive Science, 21*(3), 305-350.

Trafton, J. G., Kirschenbaum, S. S., Tsui, T. L., Miyamoto, R. T., Ballas, J. A., & Raymond, P. D. (2000). Turning pictures into numbers: Extracting and generating information from complex visualizations. *International Journal of Human Computer Studies, 53*(5), 827-850.

Trafton, J. G., Trickett, S. B., & Mintz, F. E. (2001). Overlaying images: Spatial transformations of complex visualizations. Paper to be presented at Model-Based Reasoning '01, in Pavio, Italy.

Trickett, S. B., Trafton, J. G., & Schunn, C. D. (2000). Blobs, dipsy-doodles, and other funky things: Framework anomalies in exploratory data analysis. In *The Proceedings of the 22nd Annual Conference of the Cognitive Science Society*(pp. 965-970). Hillsdale, NJ: Erlbaum.

White, B. (1993). ThinkerTools: Causal models, conceptual change, and science education. *Cognition & Instruction, 10*(1), 1-100.